CORE: Comparable, Open, Reliable, Extensible Software: An Experience Report of Four Flux Balance Analysis Tools

Alexis L. Marsh
almarsh@iastate.edu
Dept. Computer Science
Iowa State University
Ames, IA, USA

Myra B. Cohen
mcohen@iastate.edu
Dept. Computer Science
Iowa State University
Ames, IA, USA

Robert W. Cottingham
cottinghamrw@ornl.gov
Computational & Predictive Biology
Oak Ridge National Laboratory
Oak Ridge, TN, USA

ABSTRACT

Computational tools have become increasingly important to the advancement of biological research. Despite the existence of data-sharing principles such as FAIR, there has been little attention paid to ensuring that computational tools provide a platform to compare experimental results and data across different studies. Tools often are lightly documented, non-extensible, and provide different levels of accuracy of their representation. This lack of standardization in biological research reduces the potential power for new insight and discovery and makes it hard for biologists to experiment, compare, and trust results between different studies. In this poster we present our experience using four tools that perform flux balance analysis, on a set of different metabolic models. We frame our work around a proposed principle, akin to FAIR, aimed at bioinformatics tools. We call this CORE (Comparable, Open, Reliable and Extensible), and find the biggest challenges to be with comparability between tools. We also find that while the tools are all open source, without a deep understanding of the code base, they have insufficient openness. We needed to reach out to developers to resolve many of our questions, and we were still left with unexpected behaviors. We present lessons learned as a path to future improvement using the CORE principles.

KEYWORDS

systems biology, bioinformatics, flux balance analysis

1 OVERVIEW

Computational tools play a large role in scientific research, and biology is no exception. Bioinformatics tools are used to make discoveries and for decision making regarding the direction of inquiry for experimentation. They are used for prediction, exploration, simulation and data inference. They are becoming almost as important as the laboratory experiments themselves. Hence, the ability to trust these tools, and to carefully replicate results both within and between tools has become paramount.

To confidently use computational tools in research, scientists should expect to be able to readily compare, inspect and understand, trust their reliability, and to easily extend their functionality. These are fundamental precepts in order for computational tools to adhere to scientific principles, and to be used to direct research and reach scientific conclusions. Recent initiatives in bioinformatics have led to the FAIR principles (Findable, Accessible, Interoperable, and Reusable). However, FAIR's focus has primarily been on data, leaving a lack of grounding principles for the builder of computational software tools. Some more recent work has also proposed FAIR for software [1, 2] and even for scientific workflows [3], however these approaches try to directly map FAIR to software, which retains a data-centric view. We argue this is insufficient for computational tools and a new approach is needed.

In this poster, we present our experience with a set of computational tools that all perform the same core function, Flux Balance Analysis (FBA). Each of these tools is widely used and open source. And all work relatively well *out of the box*. Our original goal was to build a software testing framework, to validate a class of computational biology algorithms. However, as we discovered, this will be non-trivial. What we learned is that each tool makes its own set of assumptions about biology, supports slightly different workflows, and has a different focus and terminology. Despite all supporting the same model inputs, we found that the resulting models can vary in the number of elements, essential elements are not the same, and the ability to modify an organism's environment is bespoke. This has led us to propose a new principle, we call CORE, which focuses on computational tools. The CORE principles expect tools to be Comparable, Open, Reliable and Extensible.

ACKNOWLEDGMENTS

This work was supported in part by NSF award 1909688 and the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award DE-AC05-00OR22725.

REFERENCES

[1] Michelle Barker, Neil P Chue Hong, Daniel S Katz, Anna-Lena Lamprecht, Carlos Martinez-Ortiz, Fotis Psomopoulos, Jennifer Harrow, Leyla Jael Castro, Morane Gruenpeter, Paula Andrea Martinez, et al. 2022. Introducing the FAIR Principles for research software. *Scientific Data* 9, 1 (2022), 622.

[2] Anna-Lena Lamprecht, Leyla Garcia, Mateusz Kuzak, Carlos Martinez, Ricardo Arcila, Eva Martin Del Pico, Victoria Dominguez Del Angel, Stephanie Van De Sandt, Jon Ison, Paula Andrea Martinez, et al. 2020. Towards FAIR principles for research software. *Data Science* 3, 1 (2020), 37–59.

[3] Matthew Wolf, Jeremy Logan, Kshitij Mehta, Daniel Jacobson, Mikaela Cashman, Angelica M. Walker, Greg Eisenhauer, Patrick Widener, and Ashley Cliff. 2021. Reusability First: Toward FAIR Workflows. In *2021 IEEE International Conference on Cluster Computing (CLUSTER)*. 444–455. https://doi.org/10.1109/Cluster48925.2021.00053

BCB '23, September 3–6, 2023, Houston, TX, USA

ACM ISBN 979-8-4007-0126-9/23/09.
https://doi.org/10.1145/3584371.3613005